An in-depth analysis on the use of long short-term memory networks to predict incidence and prevalence of Covid-19 in Latin America

Bruna Lobato Barreira
brunalyuger@gmail.com
Escola Politécnica, Universidade de São Paulo
São Paulo, Brazil

Roberto Fray da Silva
roberto.fray.silva@gmail.com
Escola Politécnica, Universidade de São Paulo
São Paulo, Brazil

Carlos Eduardo Cugnasca
carlos.cugnasca@usp.br
Escola Politécnica, Universidade de São Paulo
São Paulo, Brazil

ABSTRACT

The use of machine learning techniques, especially deep learning, could improve the predictions of the currently used epidemiological models for predicting Covid-19 in the short term. This information is essential for better decision making and to reduce the impacts of the disease spread in different countries. We explored the use of support vector regression (SVR) and long short-term memory networks (LSTM), the state of the art neural network architecture for time series analysis, to predict the daily incidence and prevalence for nine countries in Latin America. Our methodology and the models used can be replicated in other countries. Our main findings were: (i) there is no single best model or best hyperparameters configuration for all countries and targets; (ii) the LSTM showed an average MAE that was around 50% lower for incidence and 20% lower for prevalence when considering all countries; (iii) the LSTM showed better results for predicting incidence for most countries (Argentina, Bolivia, Brazil, Guatemala, and Haiti); (iv) the SVR showed better results for predicting prevalence for most countries (Argentina, Bolivia, Colombia, Cuba, Guatemala, and Haiti); and (v) for Brazil, the LSTM provided better results for both targets, with an MAE that was 68% lower for incidence and 73% lower for prevalence.

CCS CONCEPTS

• **Computing methodologies** → **Supervised learning by regression**; **Neural networks**; • **Mathematics of computing** → **Time series analysis**.

KEYWORDS

Long Short-Term Memory Networks, Covid-19, Disease Prediction

ACM Reference Format:
Bruna Lobato Barreira, Roberto Fray da Silva, and Carlos Eduardo Cugnasca. 2020. An in-depth analysis on the use of long short-term memory networks to predict incidence and prevalence of Covid-19 in Latin America. In *Berlin '20: International Conference on Frontiers of Artificial Intelligence and Machine Learning, September 16–18, Berlin, DE.* ACM, New York, NY, USA, 6 pages.

Berlin '20, Sept. 16–18, 2020, Berlin.

1 INTRODUCTION

The Covid-19 pandemics generated a significant increase in the awareness of the problems related to the spread of diseases. Pandemics can be understood as diseases that spread through vast regions and many people in a short period. This results in a series of problems due to the sudden peak demand for healthcare services, especially in intensive care units [1]. In less than 2 months, the Covid-19 pandemics spread to almost all countries [2].

Decision-makers on different levels are faced with problems related to: patient allocation, resource allocation, the adoption of non-pharmaceutical interventions (such as quarantine and lockdown), and research and development of vaccines and drugs that may reduce the impact of the disease [1] [3]. It is important to note that the dynamics of the spread of the disease are still unknown. Also, the development of vaccines that could be launched in the market may take up to 18 months [1].

The most used methods to predict the disease spread are epidemiological models, such as the Susceptible Infected Recovered (SIR) and its variations, which contain assumptions related to the dynamics of the disease [4]. Those models are crucial for long-term predictions of the evolution of the disease and how it may impact a country's population on different scenarios [5]. Nevertheless, their results depend on the assumptions adopted, which are heavily influenced by the complexity of the disease's evolution [5].

An alternative to epidemiological models is the use of time series analysis (TSA) for predicting important features for decision-making, such as the daily number of new cases, new deaths, death rate, prevalence, incidence, among others. TSA methods rely on the use of the autocorrelation property of the dataset to predict new points. Several papers in the literature use TSA for predicting diseases, such as for dengue [6], influenza [3, 7, 8], pertussis [9], among others [10]. TSA models are data-driven and well suited for trend and seasonality detection. Therefore, these can complement the results obtained by epidemiological models for the short term.

The machine learning models, especially deep learning, are state of the art for TSA [11]. One of their advantages is that they do not depend on prior assumptions about the phenomenon they are modeling [12]. Therefore, it is possible to predict essential health indicators related to a disease without knowing the diseases' dynamics. The long short-term memory network (LSTM) is state of the art for several disease prediction problems, including influenza outbreaks [3, 8].

The main objective of this research is to evaluate the use of LSTM and of SVR for predicting the incidence and prevalence of

Covid-19 in nine countries in Latin America: Argentina, Bolivia, Brazil, Colombia, Cuba, Guatemala, Haiti, Mexico, and Peru. We will also evaluate several hyperparameters and how they impact the models' results.

We evaluated two main research questions: (i) does the LSTM provide better results than SVR for any of the targets, considering all countries?; and (ii) which model best predicts the targets for each of the evaluated countries?. The answers to those questions and the in-depth LSTM hyperparameters analysis in multiple countries are the main contributions of this paper. To the best of our knowledge, this is one of the first attempts to evaluate the use of LSTM for evaluating Covid-19 evolution in different countries in Latin America.

This research is structured as follows: Section 2 contains the fundamental concepts related to the models implemented and related works using them for diseases prediction; Section 3 describes the methodology used; Section 4 contains the analysis of the main results of both models, as well as an exploratory data analysis; Section 5 contains a discussion of how to adapt the analysis for other countries and the main limitations of the research; and Section 6 concludes the paper, suggesting future works.

2 MACHINE LEARNING FOR DISEASES PREDICTION

There is a vast literature on the use of machine learning for the healthcare domain. For an extensive review of the different uses of deep learning for healthcare, we refer the reader to the research by [12]. Specifically related to TSA in disease prediction, we refer the reader to the research by [10].

The work by [13] proposed the SVR algorithm, an adaptation of the support vector machine for regression purposes. The SVR objective is to predict fitting the observations very close to the boundary defined by the kernel within the space from the boundary defined by the epsilon parameter. This method demands less data for pattern recognition than the LSTM. This is important because, as observed by [14], the Covid-19 pandemics datasets are considered small compared to other problems.

The SVR is used in several pieces of research for disease predictions, such as the works on predicting influenza outbreaks by [7] and [3]. According to [15], it is one of the most widely used machine learning methods, especially for classification purposes (in this case, it is denominated support vector machine).

The LSTM was proposed by [16], and it can be defined as a feedforward neural network that uses its output as a recurrent input to learn from data with past values. The idea of recurrence was already studied, but it suffered from the vanishing problem. The LSTM solved this problem and has been proven to provide significant improvements over other models for TSA.

To the best of our knowledge, the work by [5] was the first to use an LSTM for forecasting the transmission of Covid-19. Those authors have focused on two main objectives: predicting the peak of the disease and predict the ending point of the disease outbreak. Our research has a different focus: improving the prediction of important health indicators in the short term, seeking to improve decision-making. Therefore, both pieces of research are complementary to decision-makers.

The work of [3] implemented several machine learning models to evaluate the prediction of influenza outbreaks in Syria to improve decision making. The models implemented were: linear model, SVR, gradient boosting, random forest, and LSTM. The authors concluded that the LSTM obtained the best results, with a MAPE of 3.52% and an RMSE of 0.01662. The research by [7] compared several models to predict weekly influenza outbreaks in the USA: SVR, random forest, gradient boosting, ARIMA, multilayer perceptron, and LSTM. They concluded that the LSTM obtained the best results for MAPE (5.4%) and RMSE (0.00210).

In the next section, we will describe the steps of the methodology used in this research.

3 METHODOLOGY

The methodology used in this paper consisted of six steps:

1. Data gathering from official databases for the period from 03/04/2020 until 07/03/2020, for the most impacted countries in Latin America with a population of more than 10 million people: Argentina (AR), Bolivia (BO), Brazil (BR), Colombia (CO), Cuba (CU), Guatemala (GU), Haiti (HA), Mexico (MX), and Peru (PE). We did not use data from Ecuador, Chile, and Venezuela, as the datasets contained numbers that were considerably different from the other countries, maybe indicating low-quality data. We used data gathered by the Our World in Data[1] website, which collects data from the European Center for Disease Prevention and Control (ECDC). The dataset contained the following features for each country: the total number of cases, the total number of deaths, and the number of new cases and new deaths in the day. We then collected the number of recovered patients from the Johns Hopkins University CSSE database [2] [2];

2. Data preprocessing. We created one dataset for each country. Then, we preprocessed the datasets to remove all missing data and identify possible outliers and calculated the prevalence and incidence. Lastly, we separated the datasets into three subsets: 80% of the full dataset for training and validation (using the blocking time-series cross-validation method with 5 splits), and 20% of the full dataset for testing;

3. Exploratory data analysis. We conducted an exploratory data analysis for each dataset to evaluate the autocorrelation in the data using the Augmented Dickey-Fuller test and analyzed each feature's behavior for each dataset;

4. Implementation of the SVR and hyperparameters analysis. The SVR was implemented considering all the features as inputs. We implemented one model for each target (incidence and prevalence), for each of the countries. We used the train and validation subsets of each dataset for hyperparameters analysis and final model training, and the test subset for model evaluation. The hyperparameters evaluated were: kernel, error penalty param, and epsilon. We used a grid search to find the best models and the mean absolute error (MAE) of the predictions for evaluation;

5. Implementation of the LSTM model and hyperparameters analysis. The same methodology of Step 4 was used to implement the LSTMs. The hyperparameters evaluated were: number of neurons in the LSTM layer, batch size, and number of training

[1] https://ourworldindata.org/coronavirus-source-data
[2] https://github.com/CSSEGISandData/COVID-19

epochs. We used one LSTM layer with the hyperbolic tangent (tanh) activation function, the Adam optimizer, and the MAE as a loss function;

6. Models comparison. The final models from Steps 4 and 5 were evaluated for each dataset and target to find the models that provided the best results, based on the MAE on the test subsets.

We used the following Python libraries: Pandas, Statsmodels, Scikit-Learn, Keras, TensorFlow, Matplotlib, NumPy, and Seaborn. We implemented the models using a Google Colab GPU. A total of 1680 SVR and 2000 LSTM models were implemented. The datasets and the code are available on an open Github repository[3].

4 RESULTS

This section describes the main results of the research and is divided into four subsections: 4.1 contains an exploratory analysis of the datasets; 4.2 describes the analysis of the results of the SVR implementations; 4.3 contains an in-depth analysis of the LSTM implementations, with focus on four countries: Argentina, Brazil, Colombia and Mexico; and 4.4 contains the results of the comparison of the models.

4.1 Exploratory data analysis

Figure 1 illustrates the incidence and prevalence of all countries. As the increase in the total number of cases shows exponential trends for all countries, the data was non-stationary for both incidence and prevalence. It is important to note that Cuba, Colombia, and Guatemala showed an incidence from 2 to 4x smaller than the other countries. Peru presented erratic behavior without a clear trend. Brazil, Colombia, and Guatemala showed a trend similar to an exponential curve.

Argentina displayed an increasing trend, but still slower than an exponential curve. This may be due to the different spreads of the disease in the different countries, the quality of the health services, and the data provided (especially considering occurrences of sub notification and lack of testing). Lastly, Cuba and Haiti showed a more linear trend. This may be due to the plans adopted by those countries or by lack of testing and sub notification.

Regarding the prevalence, it is important to note that the overall trends for the countries are quite similar to those observed for incidence, except for Brazil, Haiti, Mexico, and Peru. For Brazil, the trend is upward, but it showed several fluctuations. For Haiti, a plateau seems to have been reached, which may be related to a lack of testing. For Mexico, a slowly increasing trend can be observed. Lastly, for Peru, there was a fast increase in the number of cases in May, followed by a plateau in June. We believe that this is due to a lack of data quality and a lack of testing.

In the next subsection, we will explore the results of the SVR implementations.

4.2 SVR results

In this section, we analyze the SVR implementations' results. Table 1 contains the MAE on the test subset for each country, together with the hyperparameters that presented the best results.

[3]https://github.com/rfsilva1/covid19lstm/

Table 1: SVR hyperparameters values and MAE on the test subset for each country.

Country	Incidence	Prevalence
AR	K:rbf; C:1; E:0.01 / **3.509**	K:lin; C:10; E:0.001 / **2.074**
BO	K:rbf; C:20; E:0.01 / **4.665**	K:rbf; C:5; E:0.001 / **15.342**
BR	K:lin; C:50; E:0.001 / **10.396**	K:lin; C:10; E:0.001 / **66.479**
CO	K:lin; C:0.1; E:0.1 / **1.052**	K:lin; C:20; E:0.001 / **11.905**
CU	K:poly; C:0.1; E:0.1 / **0.034**	K:lin; C:10; E:0.001 / **0.310**
GU	K:sigd; C:0.1; E:0.001 / **1.708**	K:lin; C:50; E:0.001 / **1.794**
HA	K:poly; C:5; E:0.1 / **4.982**	K:lin; C:50; E:0.001 / **3.651**
MX	K:lin; C:1; E:0.1 / **0.439**	K:lin; C:1; E:0.1 / **2.165**
PE	K:rbf; C:1; E:0.1 / **1.771**	K:lin; C:20; E:0.001 / **41.556**

Legend: K: kernel, C: penalty parameter, E: epsilon.

First, it is essential to note that, as the countries' datasets presented different trends and values, the models' best hyperparameters were different. Nevertheless, it is possible to observe that: (i) the RBF is the best kernel for most countries for predicting incidence; (ii) the linear kernel is the best one to predict prevalence for most countries, except for Bolivia; (iii) the values for C tended to be smaller for predicting incidence; (iv) the epsilon for predicting incidence tended to be higher for all countries.

The countries that presented the lowest MAE for incidence were: Cuba, Mexico, Colombia, and Guatemala. Nevertheless, it is important to observe that the MAE was considerably higher for the following countries: Brazil, Haiti, Bolivia, and Argentina. One of the main reasons for explaining these observations is the significant differences in the features' values on the train, validation, and test subsets. This may be due to the faster disease spread with time.

For prevalence, the countries that presented the lowest MAE were: Cuba, Guatemala, Argentina, and Mexico. The countries that presented the highest MAE were: Brazil, Peru, Bolivia, and Colombia.

Those results can be explained mainly by three factors: (i) the considerable difference between the prevalence values on the train, validation, and test subsets; (ii) to possible data quality problems on the original data; and (iii) the considerably different trends and patterns presented on the prevalence for those countries, as was described in the analysis in subsection 4.1. All these factors increase the difficulty of the SVR to detect relevant patterns in the datasets.

Although the SVR is easier to implement, it has some disadvantages in relation to the LSTM. The first is that the SVR does not capture the data structure as a deep neural network does, so it demands that features should be created for this model. This is a problem as there may be important features that we are not aware of, as the disease is not as well studied as other diseases. The second disadvantage is that it does not consider the autocorrelation of the data on its model. Therefore, it is expected that it will provide worse results for data that show a clear trend or seasonality.

In the next section, we will conduct an in-depth analysis of the LSTM implementations.

4.3 LSTM results

The architecture of the LSTM model implemented was inspired both by the literature [5, 8, 11] and by preliminary experiments

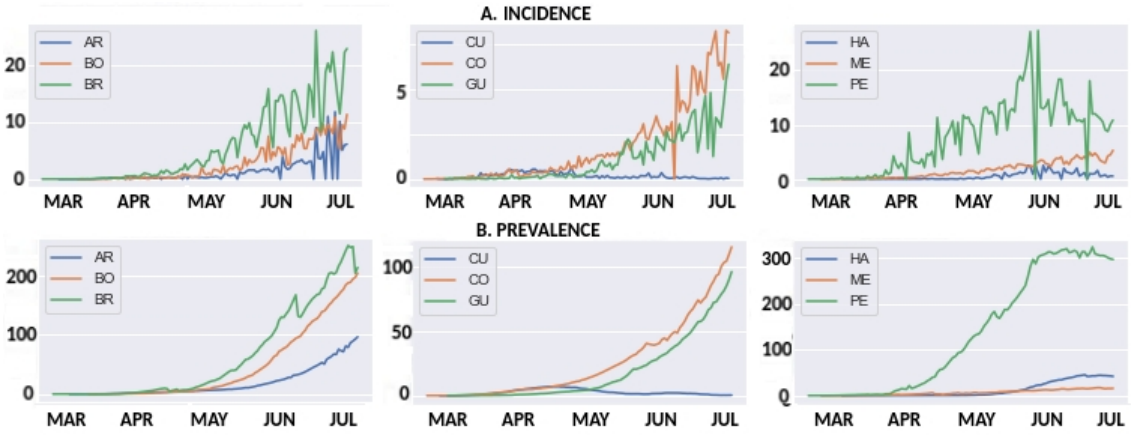

Figure 1: Incidence and prevalence of all the analyzed countries.

Table 2: LSTM hyperparameters values and MAE on the test subsets for each country, in bold.

Country	Incidence	Prevalence
AR	B:8; N:50; E:100 / **2.148**	B:2; N:500; E:50 / **11.712**
BO	B:16; N:200; E:20 / **1.181**	B:32; N:500; E:50 / **17.337**
BR	B:8; N:500; E:20 / **3.357**	B:4; N:50; E:200 / **17.933**
CO	B:32; N:100; E:20 / **2.355**	B:32; N:200; E:100 / **27.087**
CU	B:8; N:10; E:200 / **0.063**	B:8; N:500; E:200 / **0.702**
GU	B:32; N:10; E:200 / **1.327**	B:16, N:50; E:200 / **13.936**
HA	B:32; N:50; E:100 / **0.593**	B:2; N:50; E:100 / **8.366**
MX	B:16; N:500; E:50 / **0.556**	B:32; N:500; E:20 / **0.799**
PE	B:8; N:10; E:200 / **3.676**	B:2; N:100; E:200 / **14.235**

Legend: B: batch size, N: number of neurons, E: number of epochs.

conducted with the datasets. It contained four layers: (i) an input layer, which receives each data point with its features scaled; (ii) one LSTM layer; (iii) one flatten layer; and (iv) one dense layer to provide the final output. It is essential to observe that adding additional LSTM layers did not improve the models' results in our preliminary experiments. Table 2 contains the hyperparameters that presented the best results for the LSTM implementation for the different countries and the MAE for the test subsets.

First, it is possible to observe that there was no single value for any of the hyperparameters that provided the best results for all datasets. This can be explained mainly because the distribution of values for each feature varied considerably among datasets. Therefore, we can conclude that it is crucial to conduct experiments on hyperparameter tuning for each country to evaluate the evolution of Covid-19.

The countries that presented the lowest MAE on the test subset for the incidence target were: Cuba, Mexico, Haiti, and Bolivia. For the prevalence target, these were: Cuba, Mexico, Haiti, and Argentina.

The significant difference between the MAE obtained on the different countries and targets may be due to several reasons, such as: (i) the difficulty of the model in identifying patterns in the data; (ii) the quality of the data on the different datasets; (iii) the impact of the different measures adopted to deal with the disease spread on the different countries (such as quarantines and lockdown); and (iv) the quality of the health services, which impacts on the recovery rate of the severely infected patients, among others.

It is important to observe that the errors for both Cuba and Mexico were much smaller than for the other countries. Some of the reasons may be: (i) the values for incidence and prevalence were considerably lower for those countries; and (ii) the LSTM model may have identified better patterns on those datasets. Experiments with more countries are needed to explore this observation further. Another significant result is that the MAE for prevalence was considerably higher than for incidence for all countries (except for Mexico, which showed a prevalence that was 1.44 times higher than its incidence).

Analyzing the values of the hyperparameters of the models that showed the best results for each country, it is possible to observe that, for incidence, models with higher batch size (8, 16, or 32) and lower number of neurons (100 or less) tend to have better results. Around 56% of the models have 100 or more epochs.

The results for prevalence were similar, except for the number of neurons. For this target, 56% of the models had 200 or more neurons. Although the values of the datasets' features varied considerably, we believe this is an indication that those ranges of values should present good results for the targets in other countries with similar characteristics. Further experiments are needed to understand this relationship better.

We conducted an in-depth analysis of Argentina, Brazil, Colombia, and Mexico since they are the most populous countries in Latin America. Table 3 illustrates the analysis of the hyperparameters for the best models for each of those countries for both targets.

The results in this table represent, for each target and hyperparameter, how much the MAE would be reduced if the best value of the hyperparameter was chosen in relation to the worst value. The number of neurons was the most important hyperparameter to be tuned. It was the one that impacted the most each of the countries analyzed individually for both targets. For incidence, choosing it correctly could improve the results around 30%. For prevalence, it

Table 3: Improvements that could be attained by choosing the best hyperparameter value in relation to the worst one for the models implemented. The best results for each country and target are in bold.

Target	Hyperparameter	AR	BR	CO	MX
Incidence	B	**33%**	21%	12%	10%
	NN	**32%**	**33%**	**26%**	**32%**
	Ep	24%	25%	9%	20%
Prevalence	B	28%	16%	27%	**32%**
	NN	**57%**	**33%**	**71%**	**34%**
	Ep	26%	11%	52%	**33%**

Legend: B: batch size, NN: number of neurons, Ep: number of epochs.

Table 4: MAE on the test subset per country for both targets. The best results for each country and target are in bold.

Country	Incidence			Prevalence		
	SVR	LSTM	Diff.	SVR	LSTM	Diff
AR	3.509	**2.148**	**39%**	**2.074**	11.712	-465%
BO	4.665	**1.181**	**75%**	**15.342**	17.337	-13%
BR	10.396	**3.357**	**68%**	66.479	**17.933**	**73%**
CO	**1.053**	2.355	-124%	**11.905**	27.087	-128%
CU	**0.034**	0.063	-86%	**0.310**	0.702	-126%
GU	1.708	**1.327**	**22%**	**1.794**	13.936	-677%
HA	4.982	**0.593**	**88%**	**3.651**	8.366	-129%
MX	**0.439**	0.556	-26%	2.165	**0.799**	**63%**
PE	**1.771**	3.676	-108%	41.556	**14.235**	**66%**
Average	3.173	**1.695**	**47%**	16.142	**12.456**	**23%**

could provide an improvement of up to 70%, as was observed in the Colombia dataset. This result is in line with the literature, in which the hyperparameters of the LSTM layer should be the most impactful ones.

The number of epochs was the second most important hyperparameter for predicting prevalence and, even though it was not the most impactful in any of the individual countries for predicting incidence, its average across the countries (20%) was higher than that of the batch size (19%) for that target. The number of epochs is directly related to how much the neural network will be able to learn and under and overfitting. Seeking the optimal number of epochs is a critical aspect to improve the model's pattern recognition. Lastly, the batch size was the least important of the hyperparameters evaluated for those four countries. In the next section, we will compare all the models implemented.

4.4 Final models comparison

This section will compare the results of the best SVR and LSTMs implemented on each dataset. Table 4 contains the results for both targets, considering the MAE on the test subsets and the difference between the LSTM and the SVR MAE for each country.

The average MAE considering all countries is lower for the LSTM, for both targets. For incidence, the average MAE for the LSTM was around 50% lower than for the SVR. For prevalence, it was around 20% lower. In the case of incidence, this is because the SVR showed MAE values that were significantly higher for Haiti (8.4x higher), Bolivia (4x higher), and Brazil (3x higher). In the case of prevalence, the SVR showed a higher MAE for Brazil (3.7x higher), Peru (around 3x higher), and Mexico (2.7x higher).

We believe that one of the main reasons for those results is that the LSTM architecture takes into account the temporal correlation of the data, which allows it to identify better time-related patterns (such as the increasing trend in the features of the dataset). Also, it is crucial to observe that, as explored in the literature, the LSTM may also capture more complex patterns on the data [11, 12].

Another interesting result was that, in general, the LSTM presented better results for incidence (5 out of the 9 datasets had a lower MAE with the LSTM), while the SVR obtained better results for the prevalence (6 out of the 9 datasets had a lower MAE with the SVR). This may be due to the different characteristics of the datasets, as already explored in subsection 4.1.

As the LSTM considers past data to make its predictions, it may be more negatively affected than the SVR when erratic patterns are observed, as was the prevalence on the different datasets. As seen in subsection 4.1, the incidence had a more continuous trend among the datasets, which may explain the considerably better results of the LSTM for predicting this target.

Answering the first research question of this research, we can infer that the LSTM provides better results than the SVR on predicting incidence, not predicting prevalence. Therefore, we would recommend using it to predict this target in different countries. Nevertheless, it is important to observe that the SVR must still be used as a baseline, as it presents better results for some countries.

Answering the second question of this research, the countries in which the incidence was improved by using the LSTM were: Argentina (MAE: 2.1476, a 40% improvement), Bolivia (MAE: 1.1806, a 75% improvement), Brazil (MAE: 3.3572, a 68% improvement), Guatemala (MAE: 1.3269, a 22% improvement), and Haiti (MAE: 0.5927, a 88% improvement).

In the case of prevalence, the following countries observed a lower MAE with the use of the LSTM: Brazil (MAE: 17.9331, a 73% improvement), Mexico (MAE: 0.7991, a 63% improvement), and Peru (MAE: 14.2353, a 66% improvement). For the prediction of incidence for Colombia, Cuba, Mexico, and Peru, the SVR presented better results. This was also observed for the prevalence of Argentina, Bolivia, Colombia, Cuba, Guatemala, and Haiti.

An important observation is that, for the Brazilian dataset, the LSTM provided better results for both incidence (with a 68% improvement) and prevalence (with a 73% improvement). For incidence, this is in line with the previous observations. One of the reasons that may explain the observed results for prevalence in this country is that it has a more continuous growing trend than most other countries. This improves the LSTM's capacity to recognize patterns.

Lastly, two important observations are: (i) the LSTM did not improve the results for any target for Colombia and Cuba; and (ii) the LSTM resulted in considerably worse results for prevalence in Argentina (MAE for the LSTM was more than 5x higher than for the SVR) and Guatemala (MAE for the LSTM was more than 7x higher than for the SVR). Further studies are necessary to evaluate

the reasons for those results better. Nevertheless, the inconsistent trend observed in those countries' prevalence may be one of the main reasons for those results.

In the next section, we will explore how to expand the analysis to other countries and diseases.

5 DISCUSSIONS

One crucial aspect observed in our results was the variability of the models and hyperparameters values across countries. This may be the case due to several reasons that were explored throughout the paper: lack of data quality; sub notification; lack of testing; significant differences between the train, validation, and test subsets; a small dataset (which makes it harder for the LSTM to identify patterns); among others.

To further explore those results, more studies are needed with different countries. It would also be interesting to evaluate the impact of varying the distributions of the train, validation, and test sets. The use of unsupervised learning methods could also be explored to identify clusters of countries with similar characteristics based on their data or to provide feature representations to improve the machine learning models' results.

The methodology and models used can also be applied to other diseases, especially if their dynamics are partially known or unknown. In that case, especially at the beginning of the disease spread, the models implemented (together with other econometrics models such as ARIMA, SARIMA, and SARIMAX) could provide valuable information.

Lastly, as knowledge of the disease's dynamics and its spread increases, it is possible to develop better epidemiological models. In this scenario, the machine learning models implemented could provide additional information for the decision-makers.

The main limitations of this research were: (i) the small sizes of the datasets; (ii) the problems already mentioned related to data quality and sub notification; (iii) and the partially known dynamics of the disease. Nevertheless, we believe that these were fully addressed by the methodology used. In the next section, we will conclude the paper and provide recommendations for future works.

6 CONCLUSIONS AND FUTURE WORKS

In this research, we analyzed the use of LSTM and SVR for predicting the daily incidence and prevalence of Covid-19 in nine countries in Latin America. Using these models together with epidemiological models could improve governments' decision-making, by better predicting the need of hospital beds on intensive care units, monitoring disease spread, and better evaluating when possible new waves of diseases are starting.

We conducted an in-depth analysis of the LSTM results and the impacts of different hyperparameters, seeking both to: (i) contribute with the scientific community by exploring a state of the art neural network architecture applied to a relevant disease prediction problem; and (ii) provide relevant information and analyses for decision-makers regarding the short-term predictions of incidence and prevalence for Covid-19.

The main findings of the research were: (i) there is no single best model or best hyperparameters configuration for all countries and targets; (ii) the LSTM showed a lower average MAE for both targets when considering all countries; (iii) the LSTM showed better results for predicting incidence for Argentina, Bolivia, Brazil, Guatemala, and Haiti; (iv) the SVR showed better results for predicting prevalence for Argentina, Bolivia, Colombia, Cuba, Guatemala, and Haiti; and (v) the LSTM provided better results for both targets for Brazil.

Future works are related to: (i) re-evaluating the models periodically with larger datasets; (ii) implementing unsupervised machine learning models to cluster countries with similar characteristics, also considering social and economic features; (iii) evaluating the use of the models implemented together with epidemiological models to improve decision-making; and (iv) expand the analysis to more countries, to gain more insights related to how the disease spreads.

ACKNOWLEDGMENTS

This work was carried out with the support of Itaú Unibanco S.A., through the scholarship program of Programa de Bolsas Itaú (PBI), linked to the Centro de Ciência de Dados (C^2D) of Escola Politecnica da USP. We would also like to thank the National Council for Scientific and Technological Development (CNPq) for the support.

REFERENCES

[1] Neil Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. *Imperial College London*, 10:77482, 2020.

[2] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.

[3] Ali Darwish, Yasser Rahhal, and Assef Jafar. A comparative study on predicting influenza outbreaks using different feature spaces: application of influenza-like illness data from early warning alert and response system in syria. *BMC research notes*, 13(1):1–8, 2020.

[4] Howard Howie Weiss. The sir model and the foundations of public health. *Materials matematics*, pages 0001–17, 2013.

[5] Vinay Kumar Reddy Chimmula and Lei Zhang. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals*, page 109864, 2020.

[6] Sunil Bhatnagar, Vivek Lal, Shiv D Gupta, Om P Gupta, et al. Forecasting incidence of dengue in rajasthan, using time series analyses. *Indian journal of public health*, 56(4):281, 2012.

[7] Jie Zhang and Kazumitsu Nawata. A comparative study on predicting influenza outbreaks. *Bioscience trends*, 2017.

[8] Liyuan Liu, Meng Han, Yiyun Zhou, and Yan Wang. Lstm recurrent neural networks for influenza trends prediction. In *International Symposium on Bioinformatics Research and Applications*, pages 259–264. Springer, 2018.

[9] Qianglin Zeng, Dandan Li, Gui Huang, Jin Xia, Xiaoming Wang, Yamei Zhang, Wanping Tang, and Hui Zhou. Time series analysis of temporal trends in the pertussis incidence in mainland china from 2005 to 2016. *Scientific reports*, 6(1):1–8, 2016.

[10] Manliura Datilo Philemon, Zuhaimy Ismail, and Jayeola Dare. A review of epidemic forecasting using artificial neural networks. *International Journal of Epidemiologic Research*, 6(3):132–143, 2019.

[11] Luca Di Persio and Oleksandr Honchar. Artificial neural networks architectures for stock price prediction: Comparisons and applications. *International journal of circuits, systems and signal processing*, 10(2016):403–413, 2016.

[12] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6):1236–1246, 2018.

[13] Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161, 1997.

[14] Wim Naudé. Artificial intelligence against covid-19: An early review. 2020.

[15] Negin Shafaf and Hamed Malek. Applications of machine learning approaches in emergency medicine; a review article. *Archives of academic emergency medicine*, 7(1), 2019.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.